A Turing test of whether AI chatbots are behaviorally similar to humans

Year: 2024 | Citations: 203 | Authors: Qiaozhu Mei, Yutong Xie, Walter Yuan, Matthew O Jackson

Abstract

Significance As AI interacts with humans on an increasing array of tasks, it is important to understand how it behaves. Since much of AI programming is proprietary, developing methods of assessing AI by observing its behaviors is essential. We develop a Turing test to assess the behavioral and personality traits exhibited by AI. Beyond administering a personality test, we have ChatGPT variants play games that are benchmarks for assessing traits: trust, fairness, risk-aversion, altruism, and cooperation. Their behaviors fall within the distribution of behaviors of humans and exhibit patterns consistent with learning. When deviating from mean and modal human behaviors, they are more cooperative and altruistic. This is a step in developing assessments of AI as it increasingly influences human experiences.